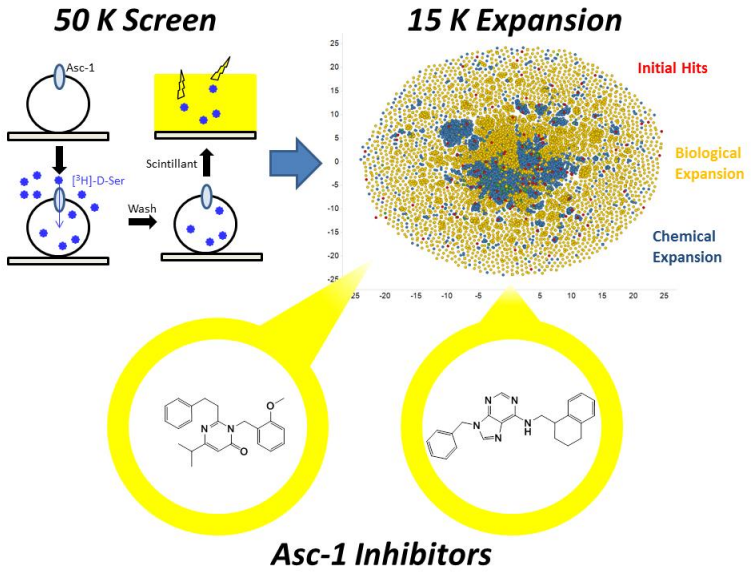
Iterative Focused Screening with Biological Fingerprints Identifies Selective Asc-1 Inhibitors Distinct from Traditional High Throughput Screening

Peter S. Kutchukian,[1] Lee Warren,[2] Brian C. Magliaro,[3] Adam Amoss,[4] Jason A. Cassaday,[4] Gregory O'Donnell,[4] Brian Squadroni,[4] Paul Zuck,[4] Danette Pascarella,[3] J. Chris Culberson,[5] Andrew J. Cooke,[6] Danielle Hurzy[6], Kelly-Ann Sondra Schlegel[6], Fiona Thomson,[2,7] Eric N. Johnson,[4,8] Victor N. Uebele,[4] Jeffrey D. Hermes,[4] and Sophie Parmentier-Batteur,[2] and Michael Finley[4,9,*]

Affiliations:
1 Modeling and Informatics, Merck & Co., Inc., MRL, Boston, MA USA
2 Neuroscience, Merck & Co., Inc., MRL, West Point, PA USA
3 Pharmacology, Merck & Co., Inc., MRL, West Point, PA USA
4 Screening and Protein Sciences, Merck & Co., Inc., MRL, North Wales, PA USA
5 Modeling and Informatics, Merck & Co., Inc., MRL, West Point, PA USA
6 Chemistry, Merck & Co., Inc., MRL, West Point, PA USA
7 current affiliation Institute of Cancer Sciences, University of Glasgow, Scotland, UK
8 current affiliation In Vitro Biology, Cellular Pharmacology, WuXi AppTec, Plainsboro, NJ
9 current affiliation Discovery Sciences, Janssen Research & Development LLC, Spring House, PA USA
* Email: finlemic@gmail.com

Graphical Abstract



Abstract

N-methyl-D-aspartate receptors (NMDARs) mediate glutamatergic signaling that is critical to cognitive processes in the central nervous system, and NMDAR hypofunction is thought to contribute to cognitive impairment observed in both schizophrenia and Alzheimer's disease. One approach to enhance the function of NMDAR is to increase the concentration of an NMDAR co-agonist, such as glycine or D-serine, in the synaptic cleft. Inhibition of alanine-serine-cysteine transporter-1 (Asc-1), the primary transporter of D-serine, is attractive because the transporter is localized to neurons in brain regions critical to cognitive function, including the hippocampus and cortical layers III and IV, and is co-localized with D-serine and NMDARs. To identify novel Asc-1 inhibitors, two different screening approaches were performed with whole-cell amino acid uptake in heterologous cells stably expressing human Asc-1: 1) a high-throughput screen (HTS) of 3 M compounds measuring ^{35}S L-cysteine uptake into cells attached to scintillation proximity assay beads in a 1536 well format and 2) an iterative focused screen (IFS) of a 45 K compound diversity set using a ^{3}H D-serine uptake assay with a liquid scintillation plate reader in a 384 well format. Critically important for both screening approaches was the implementation of counter screens to remove non-specific inhibitors of radioactive amino acid uptake. Furthermore, a 15 K compound expansion step incorporating both on- and off-target data into chemical and biological fingerprint-based models for selection of additional hits enabled the identification of novel Asc-1-selective chemical matter from the IFS that was not identified in the full-collection HTS.

Introduction

In the wake of an era where high throughput screening of large sets of compounds (2-4 M) has failed to translate into increased success in the clinic, a paradigm shift in screening has emerged, moving from an emphasis on relatively simple and low-cost high throughput screens to more biologically relevant and potentially complex screens of focused sets of compounds. In the nascent scenario, a focused screen of 10^2-10^5 compounds is often followed by an additional screen of compounds, selected based on a machine learning algorithm trained on the results from the initial library. This iterative approach ideally affords chemical starting points for a drug discovery program that perturb a more biologically relevant system, and the hope is that screening in more physiologically relevant and translatable assays will translate into greater success in the clinic. In the present report, we describe a high throughput screen (HTS) of 3 M compounds against alanine-serine-cysteine transporter-1 (Asc-1). In addition, we conducted an iterative focused screen (IFS) employing a more biologically relevant screening format. The IFS was initiated by a smaller set of diverse compounds (45 K compounds) that was followed by one step of expansion (15K compounds) selected based on machine learning. We identified tool compounds using both approaches, and offer a unique comparison of the screening paradigms.

The role of Asc-1 in regulating synaptic D-serine levels, and thereby affecting N-methyl–D-aspartate receptor (NMDAR) function[(1)], makes it a potentially attractive target for therapies relating to cognitive function[(2)]. Asc-1 knock-out studies implicate the transporter as a critical regulator of synaptic D-serine levels[(3)]. Additional studies demonstrate that increasing D-serine levels mitigates schizophrenic behaviors in animal models.[(4)] Several small molecule Asc-1-selective inhibitors have been developed with differing mechanisms of action, and recent studies have suggested that inhibition of Asc-1 may paradoxically result in reduced, rather than increased, levels of D-serine.[(5-7)] Asc-1 (encoded by SLC7A10 gene) is the "light unit" of a heterodimeric amino-acid transporter (where transport activity takes place), which also contains the "heavy unit" 4F2 cell-surface antigen heavy chain (4F2hc, encoded by the SLC3A2 gene). It is a Na^+-independent amino acid transporter with homology to the system L transporter (LAT) family.[(8, 9)] Although Asc-1 exhibits high affinity for small, neutral L-amino acids such as alanine, it also has high affinity for D-amino acids, including D-serine,[(8-10)] a co-agonist of NMDAR.[(11)] Asc-1 is widely distributed throughout the brain, largely localized to presynaptic nerve terminals[(10)] but in significant overlap with brain regions expressing and releasing D-serine from glial cells.[(12)]

We initially sought to identify selective inhibitors of Asc-1 by developing an assay that would support a screen of our full compound collection (~3 M compounds). Although radiolabelled amino acid uptake is a well-tested method for examining transporter function including Asc-1,[(5-7)] developing a cell line and an assay detection format robust enough for a miniaturized (1536-well plate based) approach was a significant challenge. We describe a radiotracer uptake screening strategy using whole cells (heterologously expressing Asc-1) adherent to scintillation proximity assay (SPA) beads with [^{35}S]-cysteine as a surrogate for D-serine. In addition we incorporated off-target assays, including uptake in parental cells and a cytotoxicity assay, to mitigate potential false-positives that affected non-specific amino acid uptake and cell viability, respectively. Based on a subsequent comparison of the SPA assay with a more biologically relevant [^{3}H]-D-Serine uptake inhibition assay against a set of previously untested compounds, we were concerned that the assay might have missed relevant and attractive

chemical matter. This propelled us to perform an IFS on the lower throughput, but more biologically relevant assay in a 384 well format. Off-target assays were also incorporated into this approach to inform the subsequent machine learning models. The focused screen (45 K compounds) was followed up with an expansions step (15 K compounds), based on compounds selected for screening from our entire compound collection using naïve Bayesian models.

Machine learning models such as random forest or naïve Bayesian have proven to be valuable in selecting compounds for screening based on a previous focused screen.[13, 14] In many retrospective analyses of IFS, full deck screening data is used to validate models, and the hits obtained from a focused set followed by several expansion steps are compared to the total hits obtained by the full deck screen.[13, 15, 16] In practice, however, this is usually not a feasible approach, and only 1-3 steps of expansion are performed.[13] Here, we built two models on the initial 45 K focused screen, and selected 7.5 K compounds with each model. For one model we used chemical descriptors (ECFP4[17]), and for the other model we used biological descriptors (HTS fingerprints[18]). While there may be some overlap in the compounds that are identified with the two approaches, we anticipated that they would complement each other and result in a broader coverage of the active chemical matter.[14, 15, 18] In addition to the information captured in the screen (active/inactive) we also included compounds in the "inactive" category that contained chemotypes that we wanted to avoid, previously identified as undesirable by the project team during the work up of the full deck HTS screen. This approach allowed us to compare, head to head, models trained on either chemical or biological descriptors, in a prospective manner. This is the largest *prospective* comparison of chemical versus biological expansion for a single assay to date.

In the present report, we seek to address three questions: 1) After previously running and working up a full deck screen against a challenging target such as Asc-1, can a screen aimed at the same target but using a more relevant biological format yield novel chemical matter that can serve as drug leads? 2) How does expansion of an initial screen based on chemical structural fingerprints compare to expansion based on biological fingerprints? 3) How do the small molecule tools identified using a focused approach compare to the previously identified compounds from the full HTS?

Results and Discussion

Optimization of HTS SPA Assay for Asc-1 Inhibition. The first strategy to identify novel inhibitors of Asc-1 relied on the ability to screen our company's 3 M compound library in a 1536 well format. Although the measurement of D-serine would be more biologically relevant as the natural substrate of Asc-1 relevant to NMDAR modulation, this amino acid can only be labelled with the relatively weak beta emitter ^{3}H. The lower energy radioactive decay requires screening methods that are less efficient (see Methods). In contrast, a combination of a higher energy beta decay isotope (^{35}S) on a surrogate sulfur-containing amino acid, cysteine, in combination with SPA beads would potentially enable a higher-density screening format. The SPA imaging bead format (Figure 1A) was examined with whole-cells stably expressing Asc-1 and mediating uptake of [^{35}S] cysteine to achieve a sufficient assay window.[6] Initial assay development was performed in a 384 well plate format, and Asc-1-specific uptake of [^{35}S] cysteine was established with inhibition by 10 mM D-serine, relative to total uptake

determined by inhibition with 1 mM cold cysteine. Optimal SPA bead coating was determined to be poly-L-lysine rather than wheat germ agglutinin (Figure S1A). Incubation times were also explored (Figure S1B), identifying an optimal incubation time to maximize Asc-1-specific [^{35}S] cysteine uptake between 12 – 19 minutes (Figure S1B Inset). The assay window was further improved by addition of other amino acids to suppress non-Asc-1 dependent transport (Figure S1C). Although L-phenylalanine appeared to give the largest window, it also increased background and was difficult to solubilize, so L-glutamine was selected instead. Subsequent optimization steps were performed to achieve the final 10 µL per well 1536 well plate format as summarized in Figure S1D.

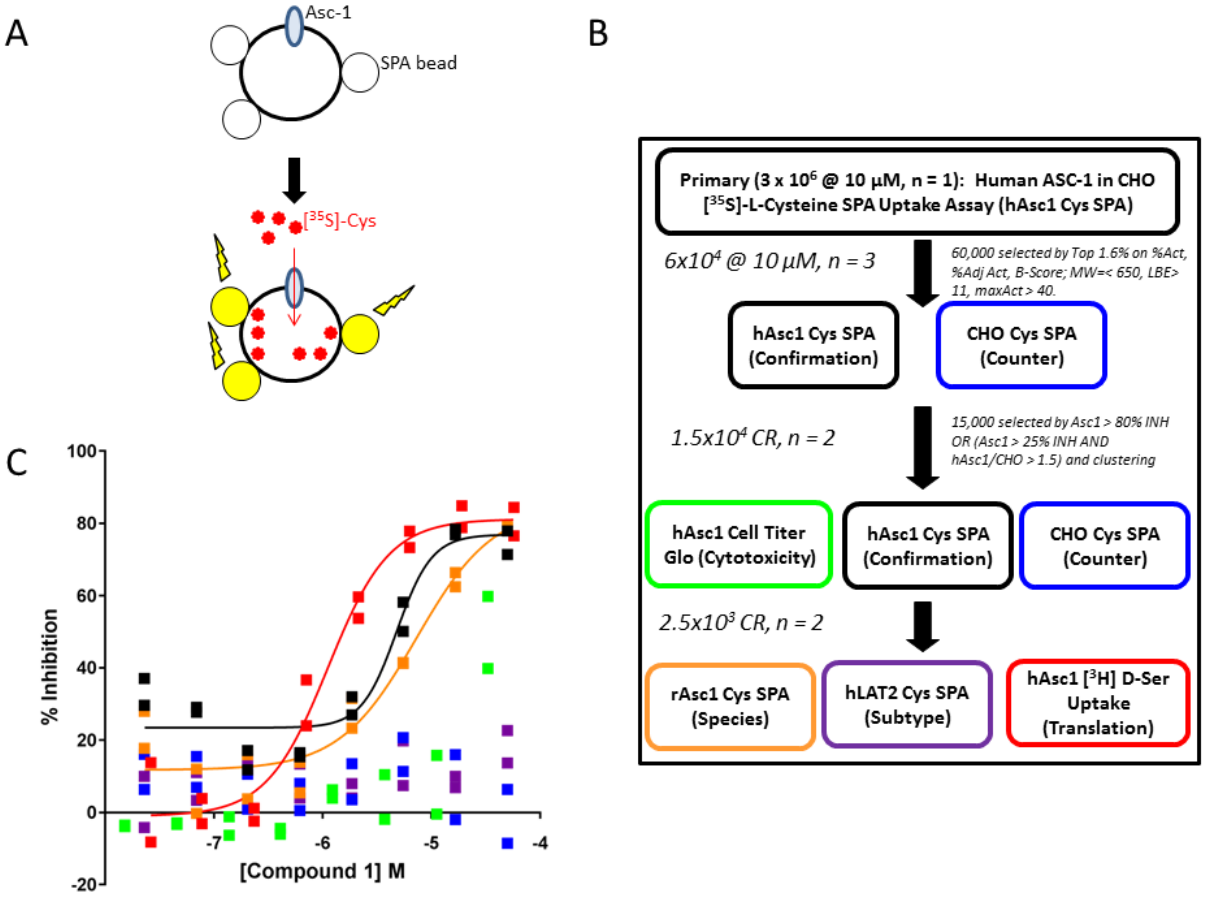


Figure 1. Whole-Cell SPA HTS A. Schematic of whole-cell SPA assay in which cells expressing Asc-1 are mixed with scintillation-containing beads to enable binding in the assay well. Upon addition of ^{35}S L-cysteine, radioactive amino acid is concentrated inside the cell, enabling energy transfer from radioactive decay to the scintillation bead which emits light that is detected by the plate imager. B. HTS screening funnel for full deck screen is shown indicating different assays by outline color and different stages: primary, confirmation/counter, concentration-response. Numbers of compounds screened at each step are indicated on the left and a brief summary of selection criteria are indicated on the right. CR indicates concentration-response performed with 8 concentrations at 3-fold dilution. C. Concentration-titration data from Compound 1 representing a desirable profile emerging from the screening funnel in B. Note that the data colors correspond to the assay outline color in B. Lines indicate best fit to Equation 1 for titrations that could be fit.

HTS SPA Assay for 3 M Compounds. High-throughput screening of the full compound collection at 10 µM final concentration (n = 1) spanned 2374 plates with an average signal-to-baseline of 2.6 ± 0.5 fold and Z' 0.40 ± 0.12 (Figure S2 A, B). Plates containing titrations of previously identified non-amino acid small molecule inhibitors of Asc-1 were tested throughout the primary screen by including these plates in runs of 100 – 120 compound collection plates (Figure S2 C, D). Monitoring these positive controls to confirm that they were also active in the assay allowed an additional qualitative assessment of assay quality beyond the intraplate positive (10 mM D-Serine) and negative (DMSO alone) controls. Hits from the screening campaign were selected based on four different measurements: % Activity, Z-score, Adjusted %Activity and B-score, with the latter two measurements reflecting corrections for systematic spatial adjustments during the assay.[19] An additional culling of the list was done to identify compounds with desirable properties (molecular weight < 650, approximate ligand binding efficiency > 11). A set of 60,000 compounds were tested in triplicate in the primary screening assay (confirmation) as well as an off-target, non-Asc-1 mediated uptake assay in CHO-K1 cells (counter screen, Fig 1B). Compounds with confirmed activity, lack of activity in the counter screen, and desirable physical chemical properties (15 K) were selected for 8 point titration testing in the Asc-1 and non-Asc-1 uptake assays as well as a cytotoxicity assay (Fig 1B). In an additional round of follow up, 2500 compounds were subjected to a 384 well plate [^{3}H]-D-Serine uptake assay, a rat Asc-1 assay to confirm activity on a species for in vivo profiling, and a human LAT-2 assay to identify compounds that broadly inhibit the system L transporter family. Figure 1C illustrates activity data from a compound (Compound **1**) with a desirable profile emerging from the screening plan illustrated in Figure 1B, exhibiting a concentration-dependent effect on Asc-1 uptake in both the [^{35}S] cysteine and [^{3}H]-D-Serine uptake assays uptake assays and no observed effect in the cytotoxicity or the CHO-K1 assay. While there were several confirmed hits / series (> 5) that were identified from the HTS SPA assay, compound **1** represents a hit from one of the optimizable series derived from this campaign (Figure 5).

HTS SPA Assay for Asc-1 Prone to False Negatives. As chemical matter from the full compound library screen was under further exploration, a small set of compounds were added to our company's screening library as part of a routine update to the collection. We took this opportunity to directly compare the [^{35}S] cysteine whole-cell SPA assay to the [^{3}H]-D-Serine uptake assay by screening the same set of 2440 compounds in both assays. Selecting the top 2% of actives from each assay (representing % Inhibition cut-offs of > 68 % and > 26 %, respectively) revealed that both assays showed a relatively small overlap in hit detection (Figure 2, green box, 9 compounds), with additional compounds that appeared to have Asc-1 inhibitory activity only detected by the [^{3}H]-D-Serine uptake assay (Figure 2B, red box, 39 compounds). Furthermore, increasing the hit selection in the [^{35}S] cysteine whole-cell SPA assay to the top 5 % (50% Inhibition cut-off) only increased the overlap to 20 compounds, leaving 28 compounds only identified in the [^{3}H]-D-Serine uptake assay. These results suggested that the [^{3}H]-D-Serine uptake assay may be identifying novel chemotypes, potentially found in 1.1 to 1.5% of the collection (this is a rough estimate extrapolating from 28 to 39 assay-unique actives of the 2440 set), that could not be detected by the SPA assay. This observation compelled us to initiate an IFS with the [^{3}H]-D-Serine uptake assay where Asc-1 uptake of the physiologically relevant NMDAR co-agonist D-

Serine could be directly assessed. Although we did not have empirical evidence that the [^{3}H]-D-Serine uptake assay was better correlated with ex vivo or in vivo Asc-1 inhibitor activity across a wide set of compounds, D-serine uptake had been employed in other prior screening efforts to identify such inhibitors[(6)].

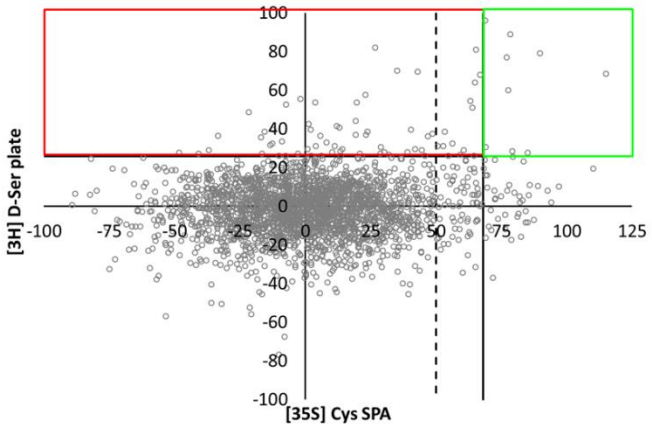


Figure 2. A plot illustrating % inhibition in the [^{35}S] cys SPA assay versus the [^{3}H] D-ser uptake assay for a set of 2440 compounds tested in both assays in parallel. The green box highlights the overlap of the top 2 % of hits in each assay, whereas the red box highlights the hits that would only be detected in the D-ser uptake assay. The vertical lines indicate the % Inhibition cut off that give the top 2% (solid) or top 5 % (dashed) of actives in the [^{35}S] cys SPA assay. Note that with either of these cut-offs there remain compounds with activity in the [^{3}H] D-ser uptake assay that would not have been identified as hits in the [^{35}S] cys SPA assay (i.e. false negatives).

Iterative Focused Screen with [^{3}H]-D-Serine Uptake Assay. We next explored an iterative focused screening strategy (Figure 3B) based on the lower-throughput but more biologically relevant [^{3}H]-D-Serine uptake assay (Figure 3A). Assay optimization was performed to adapt to a fully-automated 384 well format (Figure S3). A structurally diverse collection of ~45 K compounds was employed in the initial round of screening. Importantly, a counter screen was incorporated into the screening strategy. Initially, [^{3}H]-D-Serine uptake in a CHO-K1 line lacking Asc-1 expression was explored; however, there was insufficient signal window to utilize this assay approach (data not shown). Instead, using measurement of [^{3}H]-L-tyrosine uptake in CHO-K1 cells in a parallel 384-well, plate-based scintillation assay format exhibited sufficient assay window to eliminate false positives due to non-specific inhibition of amino acid uptake (Figure S4). In order to be considered a "hit," a compound had to demonstrate >50% [^{3}H]-D-Serine uptake inhibition, < 50% [^{3}H]-L-tyrosine uptake uptake inhibition (counterscreen), and a 20% difference between the two values (to ensure specificity). These criteria proved quite stringent, and a hit rate of 0.26% was obtained in the initial screen.

Following further confirmation of the initial hits, an expansion of hits was performed using two different modeling approaches, one based on 2D chemical fingerprints, and one based on biological

fingerprints. For both models, in addition to using the actives and inactives from the 45 K screen in training, undesirable compounds from the original full collection HTS SPA assay were also included in the "inactive" category, to avoid rediscovery of these chemotypes during the expansion step. For each model, the top scoring 7.5 K compounds were selected from our company's screening collection (see methods), totaling ~15 K compounds. Out of 7.5 K compounds selected by each method, 363 compounds (~5%) were identified by both methods. When the 15 K compounds were screened in the expansion step, there were 600 hits, a hit rate of 4.2 % (16-fold enrichment over the initial screen; same hit criteria as above). Interestingly, if the two models are considered separately, the hit rates were nearly identical (4.7 % for ECFP4, 4.6% for HTS-FP; hit rates are higher here due to the slight overlap of the selected compounds).

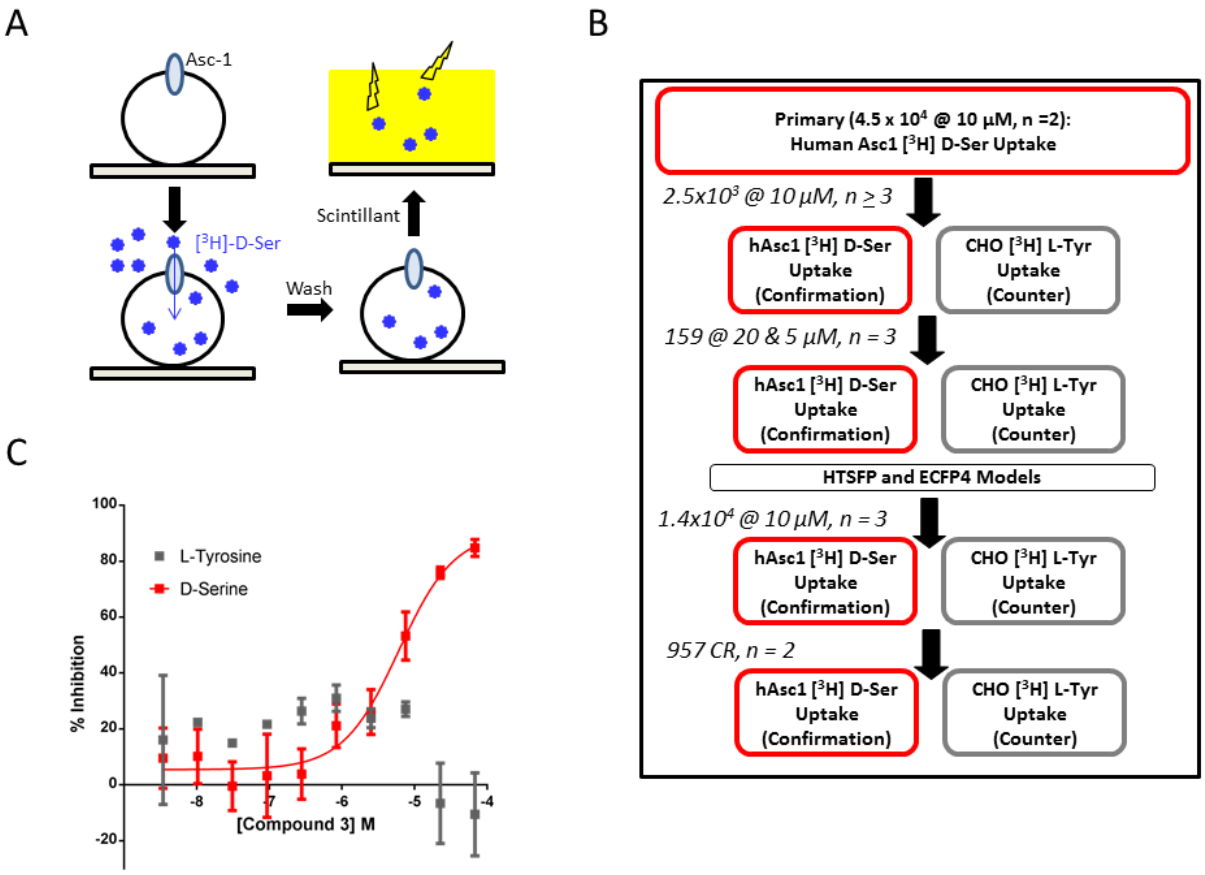


Figure 3. A. Schematic of plate-based [^{3}H] D-serine uptake, in which cells are attached to a well bottom and incubated with [^{3}H] D-serine. Again, the radioactive amino acid is concentrated inside the Asc-1 expressing cells; however, a wash step is required to remove the extracellular material. This step is followed by cell lysis and addition of scintillant to detect the weaker beta emissions from the remaining [^{3}H] D-serine. B. Screening funnel for the iterative focused screen is shown followed by chemical and biological expansion of hits (B). CR indicates concentration-response with 7 concentrations at three-fold dilution. C. Concentration-response data from Compound 3 in the Asc-1 [^{3}H] D-serine uptake and non-Asc-1 [^{3}H] L-tyrosine uptake assay. Note that the color of the data points correspond to the assay outline in B.

Biological versus Chemical-Based Expansion. We were curious how the two methods compared in terms of navigating chemical space from the initial hits. We employed a Non-Linear Map (NLM) to visually compare the chemical space occupied by the expansion sets based on chemical descriptors (blue) and biological descriptors (yellow) with the initial hits (red, Figure 4A). In this type of a representation, compounds that are aggregated near each other are structurally similar, though distances between various clusters of compounds are not meaningful. We noticed that in many cases, clusters obtained from chemical expansion were clustered together with one of the initial hits, whereas many of the biological clusters were devoid of any initial hits. This is not entirely surprising, since the biological expansion is entirely agnostic of chemical structure.

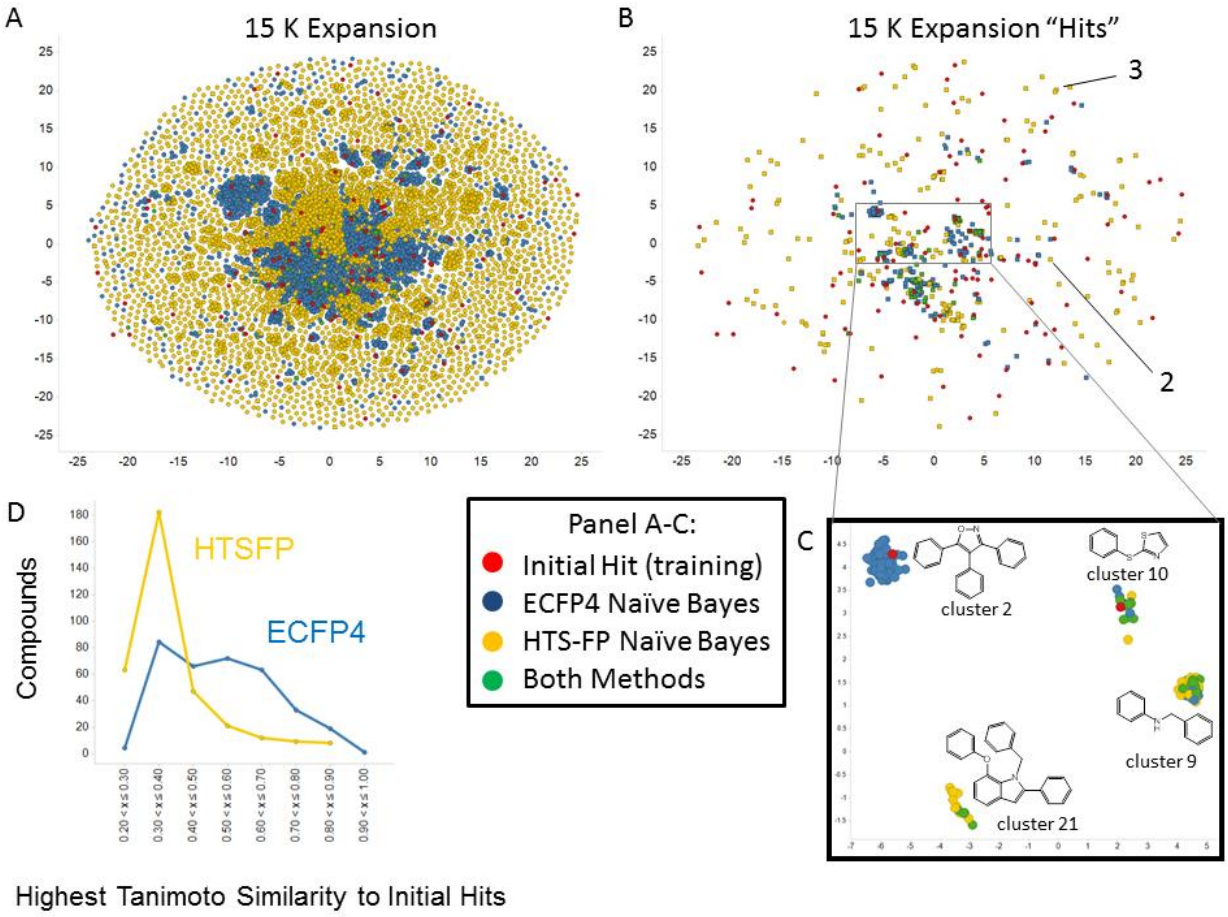


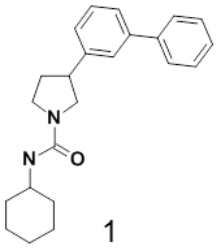
Figure 4. A: The chemical space of initial hits (red) and expansion compounds identified by chemical (blue) and biological (yellow) descriptor-based models is visualized with a Non-Linear Map (NLM). Compounds identified with both methods are in Green. Compounds that are close to each other in chemical space are clustered together on the NLM. **B**: Hits from the expansion step are visualized in chemical space. Compounds identified as interesting to follow up on by chemists are denoted as squares, and all other compounds are circles. Compounds 2 and 3 are highlighted. **C**: Four representative clusters are depicted on the NLM, with their corresponding Bemis-Murcko Scaffolds. The clusters demonstrate that similar compounds were also clustered near each other on the NLM. Some clusters were dominated by biological expansion (cluster 21) while some clusters were dominated by

chemical expansion (cluster 2) **D**: A comparison of the hits obtained through chemical expansion (ECFP4) versus biological expansion (HTSFP). Tanimoto similarity of each of the hits from the 15 K screen is compared to their nearest neighbor in the initial set of hits obtained from the 45 K screen. While hit rate were the same for each method, biological expansion identified chemical matter that is more distant from initial hits (lower Tanimoto similarity).

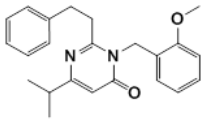
We then focused on the hits obtained from the expansion (Fig 4B). From this perspective, it is clear that we obtained diverse chemical matter to follow up on after expansion (i.e., the hits were not coming from only a few clusters). We also observed that although there was some overlap in the chemical matter that was obtained from the two methods (green, Fig 4B), the two methods were complementary and by employing both approaches, a greater diversity of chemical starting points was obtained. We highlight four of the largest clusters that were enriched in hits and their corresponding representative Bemis-Murcko scaffold (Fig 4C).[(20)] We see that clusters are indeed grouped together on the NLM. Furthermore, we can see that some clusters are dominated by chemical expansion (cluster 2), while other clusters are dominated by biological expansion (cluster 21).

To further assess the similarity of the expanded hits compared to the initial hits, we plotted the maximum similarity of the expanded hits obtained either by chemical (ECFP4) expansion or biological (HTS-FP) expansion in relation to the initial hits (Fig 4D). It is clear from this plot that although both methods had equal hit rates, biological expansion identified active chemical matter that tended to be distinct from the initial hits. We also compared the biological similarity (HTS-FP Pearson Correlation) of the expansion hits with the initial hits, revealing that in general, as might be expected, the biological similarity was greater for the expansion compounds obtained by biological (HTS-FP) expansion (Figure S5).

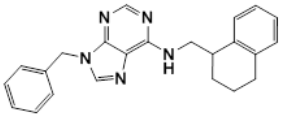
Asc-1 Tool Compounds. While several series were explored based on the expansion, two series yielded especially promising tools for Asc-1. These are represented by compound **2** and **3** (Fig 5). Of note, while there was no bias towards compounds obtained from either expansion in follow up experiments (squares in Fig 4B), both of these compounds were obtained through biological expansion of the initial hits (Fig 4B). Compound **2** was a singleton while compound **3** was in a cluster of two compounds, both of which were active and identified through biological expansion. In Figure 5, we compare their profiles to that of a representative compound obtained through the initial HTS SPA assay. It is clear that they offer distinct chemotypes for Asc-1. Furthermore, their activity in the [^{3}H] D-serine uptake Asc-1 screen is comparable to the HTS derived compound. Importantly, compounds 2 and 3, as well as a more potent analogue of Compound 1 (structure not provided), were active in a rat hippocampal neuron assay (Figure 6). Even so, other undesirable properties present in the HTS hit **1** (low solubility, high AlogP) were also present in compounds **2** and **3**, which highlights the challenging nature of identifying drug like chemical matter for the Asc-1 amino acid transporter. Despite these challenges several analogues from the Compound 1 chemical series were explored in *in vivo* models that will be part of a subsequent publication.



1



2



3

HTS Hit | HTSFP Expansion | HTSFP Expansion

Assay / Study	Compound 1	Compound 2	Compound 3
[^{3}H]-D-serine uptake Asc-1: IC_{50} (species)	2.2 µM (hum), 4.5 µM (rat)	5.2 µM (hum), 11.6 µM (rat)	5.7 µM (hum), 2.0 µM (rat)
LAT-2 IC_{50} (human)	>68 µM	>68 µM	>68 µM
CHO Tyr uptake IC_{50}	>68 µM	>68 µM	>68 µM
CellTiter-Glo Tox IC_{50}	>68 µM	>68 µM	>68 µM
ALogP98	4.8	4.5	4.6
Solubility (pH 7)	3 µM	<3 µM	<3 µM
Plasma Free fraction	0.05% (human)	0.32% (human)	0.2% (human)

Figure 5. Novel tools identified by focused screen and expansion, and their comparison to the initial HTS campaign hits.

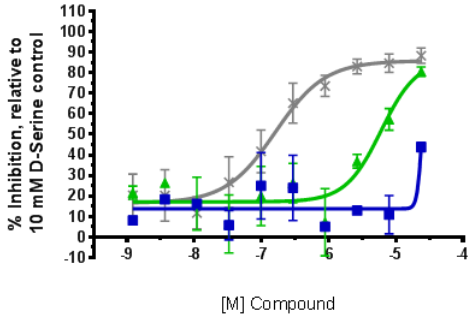


Figure 6. Compound activity in rat hippocampal neuronal [3H] D-ser uptake assay. The ability of the compounds to inhibit [^{3}H]-D-Serine uptake was evaluated in rat primary hippocampal neuron cultures. Compounds were titrated with 3-fold dilutions starting from 23.9 µM, and were added to the neurons for 10 minutes prior to adding the [^{3}H]-D-Serine. The neurons were then exposed to the [^{3}H]-D-Serine for 10 minutes to allow for uptake, followed by washes to remove the extracellular [^{3}H]-D-Serine. An optimized analog of compound 1 (grey X) showed an inflection point of inhibition at 0.1217 uM, with 88.6% inhibition at 23.9 µM (n=6), Compound 2 did not

show an inflection point up to the dose tested (>23.9 uM), with 43.4% inhibition at 23.9 uM (n=2, blue squares), and Compound 3 showed an inflection point of inhibition at 7.109 uM, with 80.3% inhibition at 23.9 uM (n=2, green triangles).

Conclusions. Our work has laid a precedent for the value in re-mining a sample collection with a more biologically relevant lower throughput screen, after a high throughput screen has already been executed on the same collection. During the IFS, we compared chemical versus biological expansion, and conclude that it is imperative that both types of models are employed in order to maximally explore chemical space during expansion steps. Through our campaigns, we have identified three Asc-1 tools that are distinctive from literature tools and can be used by research teams to further understand Asc-1 biology.

Methods

Scintillation Proximity Assay (SPA) Imaging Bead Assay: Chinese Hamster Ovary (CHO) cells stably expressing human Asc-1/4F2hc (SLC7A10/SLC3A2) were maintained in F12 media with GlutaMAX and 10 % heat inactivated fetal bovine serum (Life Technologies) with 1 mg/mL geneticin and 250 ug/mL zeocin (both from Invitrogen). Cells were harvested with TrpLE (Life Technologies) and resuspended at $2x10^6$ cells/mL in Assay Buffer (in mM): 5 KCl, 1.8 CaCl2, 1.2 MgCl2,1 KH2PO4, 10 D-glucose, 20 HEPES, 130 Choline Chloride, pH 7.25 with KOH, and 4 mM L-Glutamine (all chemicals from Sigma, except for L-glutamine from Fluka). The assay was run in a 1536 well white side/bottom, non-binding plate (Corning, No. 3865). For each well, reagents were added in the following volumes and order: 2 uL Assay Buffer, 50 nL of test compound transferred by pintool (GNF), 3 uL of resuspended human Asc-1 cells (6000 cells/well final), 3 uL of polylysine-coated SPA beads (Perkin Elmer, No. RPQ0787) at 5 mg/mL in Assay Buffer. SPA beads were maintained in suspension with a recirculating apparatus. Following a 10 min incubation at room temperature, uptake was initiated with the following additions: 1 uL Assay Buffer OR 1 uL Assay Buffer + 100 mM D-Serine (10 mM final, determine 100% inhibition), then 1 uL of ^{35}S L-cysteine (Perkin Elmer, No. NEG022T, 1075 Ci/mmol, at 25 - 35 nM, 2.5 - 3.5 nM final with cold L-cysteine added to reach 100 nM total L-cysteine). The final volume of the entire reaction was 10 uL per well. Following 13 min incubation at room temperature, the plate was spun (VSpin, Agilent) and imaged on a ViewLux (PerkinElmer, SPA Red Shift Luminescence, 2X bin, emission 613 nm).

The counter screen followed the protocol above with the following changes: a) CHO-K1 cells were used instead of human Asc-1 expressing cells, b) glutamine was NOT added to the assay buffer, c) 1 mM L-cysteine was used instead of 10 mM D-Serine to determine 100 % inhibition.

[^{3}H] D-Serine Uptake Assay: Cells (as above) were plated into 384 well, white, solid-bottom, untreated plates (Greiner, No.789163) at $5x10^5$ cells/mL in 30 uL per well (15,000 cells/well final). Cells were maintained at 37°C, 98% humidity and 5% CO_2 overnight (GNF Incubator). The following day, plates were treated as follows: 2 washes of 70 uL per well of Assay Buffer leaving 5 µL, addition of 25 µL Assay Buffer or Assay Buffer + 10 mM D-Serine, pintool (GNF) transfer of 200 nL of test compounds or DMSO vehicle. Following a 10 min incubation, 5 µL of ^{3}H D-serine solution (American Radiolabeled Chemicals,

No ART0728, 20-22 Ci/mmol, at 315 nM stock, 45 nM final) was added along with cold D-serine (315 nM) resulting in 90 nM final total D-serine in 35 uL volume per well. The plate was incubated for 20 min at room temperature. The plate was then washed 3 times, followed by addition of 50 uL per well of Microscint-20 (PerkinElmer No. 6013621). Counts per minute were determined on the TopCount (PerkinElmer) after at least one hour incubation with Microscint-20.

[^{3}H] L-tyrosine Uptake Assay: CHO-K1 cells were prepared in the same manner as described for the D-Serine uptake assay. Instead of [^{3}H] D-serine, [^{3}H] L-tyrosine was used to measure non-specific amino acid uptake into the cells and L-alanine (10 mM) was used as the positive control.

Rat Hippocampal Neuron [^{3}H] D-Serine Uptake Assay: Sodium-free Assay Buffer was prepared to reduce basal activity of sodium-dependent transporters: 130 mM Choline Chloride, 5 mM KCl, 1.8 mM $CaCl_2$, 1.2 mM $MgCl_2$, 1 mM KH_2PO_4, 10 mM D-glucose, 20 mM HEPES, pH = 7.25. D-Serine (Sigma #S4250) stocks were made fresh the day of the assay by dissolving the appropriate amount of D-Serine into assay buffer +/- DMSO as necessary. The final concentration for maximal effect was 10 mM, and the starting concentration for the D-Serine curves was also 10 mM (diluted 1:3 for 10-point titrations). [^{3}H]-D-Serine (Moravek Biochemicals #MT944) was used at 90.9 nM final concentration, prepared in assay buffer. Test compounds were titrated using acoustic dispensing on the ECHO ACCESS into ProxiPlates. 10 mM compound stock solutions in DMSO were transferred appropriately, and DMSO was back-filled to keep DMSO concentrations in all wells the same. Plates were sealed, spun at 1250 rpm for 30 seconds, and stored until use during the assay.

Cells were isolated from fresh embryonic day 18 Sprague/Dawley rat hippocampal tissue (BrainBits), plated at 30,000 cells/well into PDL-coated 384 well plates and incubated at 37°C and 5% CO_2 for ~ 2 weeks (Supporting Information). On the day of the assay, cell plates were equilibrated to room temperature. Compound plates were thawed and spun at 1250 rpm for 30 seconds. One bottle of Assay Buffer (~1 L) was warmed to 37°C, and a second bottle was put on ice. To begin the experiment, the media was removed from the cells by flicking the plate over a waste container, and the cells were washed with warmed Assay Buffer on the BioTek ELX405 Select CW cell washer (3 X 50 uL wash/aspirate cycles, leaving ~ 40 uL of buffer in each well). The titrated compounds were then diluted with assay buffer and transferred to the cell plates on the Bravo liquid handler, using a 384-tip head. Assay Buffer or 110 mM D-Serine (control wells) was added to the compound plate, mixed, and transferred at 5 uL per well from the compound plate to the cell plate. The cell plates were then centrifuged at 500 rpm for ~ 30 seconds, and allowed to incubate at room temperature for 10 minutes. [^{3}H]-D-Serine was then added to the cell plates on the Bravo liquid handler, transferring 10 uL of [^{3}H]-D-Serine in assay buffer to each well of the cell plates, using a 384-tip head. The cell plates were then centrifuged at 500 rpm for ~ 30 seconds, and incubated at room temperature for 10 minutes. [^{3}H]-D-Serine uptake was then stopped by washing the cells with cold Assay Buffer on the BioTek ELX405 Select CW cell washer (6 X 100 uL wash/aspirate cycles, leaving ~ 10 uL of buffer in each well). The plates were then flipped over and tapped firmly onto paper towels to remove any excess buffer. White bottom seals (Perkin Elmer #6005199) were applied to each plate. 50 uL of Microscint 20 (Perkin Elmer #6013621) were added to each well of the plates on the Bravo liquid handler. Clear Top seals (Perkin Elmer #6005185 or #6005250) were then applied to each plate, and the plates were allowed to shake on a plate shaker

(speed 2-3) at room temperature for at least 1 hour. The plates were then read on a TopCount (Perkin Elmer Model #C384), recording the CPMA values 1 minute/well.

Data Analysis for Assays:

Data analysis for validation plates and for concentration-response curves were performed in Prism (GraphPad) or ABASE (IDBS)

using a four parameter variable slope equation 1:

$$\% \text{ Inhibition} = IC_0 + (IC_0 - IC_{100})/(1+10^{((\text{LogIC50-X})n)}).$$

Additional screening conditions and details are provided in the Supporting Information.

Naïve Bayesian Models and Expansion:

Hit calling: 45,127 compounds were tested in the primary IFS screen, and primary hits were then tested in confirmation. To be considered a confirmed hit, compounds needed to demonstrate >50% D-serine uptake inhibition, < 50% L-tyrosine uptake inhibition (counterscreen), and there needed to be a >25% difference between D-serine and L-tyrosine uptake inhibition. Models were trained on a combination of confirmation data and chemotypes to avoid from the previous HTS as follows. To avoid rediscovering chemotypes that were previously identified as problematic by the project team (off targets, inactive, toxic, non-specific) a list of "exclude" compounds was generated (82 compounds). In order to be considered a "hit" for model training, a compound had to be considered a confirmed hit *and* not be a member of the exclude list of problematic compounds (117 compounds). Furthermore, the exclude compounds were added to the non-hit category prior to model training. Compounds that were tested in the primary screen, but were not available for confirmation were not included in the model.

ECFP4-based Naïve Bayesian Model: A Naïve Bayesian model was built using the hit calling described above, using the 'Learn Good Molecules' component in PLP with ECFP_4 as descriptors (Pipeline Pilot 8.5). The leave-one-out ROC AUC was 0.79. An expansion set of 2.7 million compounds, our company's screening deck in 2014, was then scored with the model. We removed compounds with >1 structural alert, filtered for a molecular weight 200-500, removed any compounds that were previously screened in the 45 K screen, and ensured that 30 uL material available. We then filtered for compounds that either had 0 structural alerts or a Naïve Bayes Score > 10, and then picked to top scoring 7,500 compounds.

HTS-FP-based Naïve Bayesian Model: A Naïve Bayesian (NB) model was built using the hit calling described above, using the 'Learn Good Molecules' component in PLP with internally derived HTS-fingerprints as descriptors (Pipeline Pilot 8.5). A cross validation for the NB model with a random 80% training set and 20% test set yielded a ROC AUC of 0.77. This compared favorably with a Random Forest (RF) model using the same training and test set (ROC AUC of 0.70, see SI for details on the RF model). Thus a NB model trained on the entire training set was employed for scoring. For the internal HTS-FP, each primary assay included in the fingerprint is considered as a separate descriptor, and the z-scores

are binned into 10 equipopulated bins, as implemented by the component. The fingerprint contains 344 assays, but for any given compound, only a subset of these assays contains a value. If a value is empty, it is ignored during training and testing. The HTS-FP-based Naïve Bayesian model was employed to score the same 2.7 expansion set described above, and the same filtering criteria was employed. A total of 7,500 compounds were selected, and merged with the ECFP-4 expansion list for screening.

Clustering: An in-house fragment based clustering algorithm was used, which is described in the Supporting Information.

Non-Linear Map: A Non-Linear Map was derived using the an in-house version of the Stochastic Proximity Embedding (SPE) algorithm described by Agrafiotis.[(21)]

Compound Characterization: Compound 1 (3-([1,1'-Biphenyl]-3-yl)-N-cyclohexylpyrrolidine-1-carboxamide)

3-(3-Bromophenyl)pyrrolidine (500 mg, 2.211 mmol) was dissolved in dichloromethane. Triethylamine (616 µl, 4.42 mmol) was added followed by cyclohexyl isocyanate (282 µl, 2.211 mmol). The reaction mixture was stirred at room temperature for 30 minutes. The reaction was diluted with dichloromethane, washed with water, dried, filtered and concentrated. The crude product was subjected to silica gel chromatography (0-60% ethyl acetate/hexanes) to give the intermediate 3-(3-bromophenyl)-N-cyclohexylpyrrolidine-1-carboxamide as a white solid (550 mg, 1.566 mmol). MS: m/z = 353.0 (M +1).

3-(3-Bromophenyl)-N-cyclohexylpyrrolidine-1-carboxamide (50 mg, 0.142 mmol), phenylboronic acid (17.35 mg, 0.142 mmol), bis(triphenylphosphine)palladium(II) chloride (5.00 mg, 7.12 µmol) and sodium carbonate (214 µl, 0.427 mmol) were combined in dioxane (712 µl). The reaction mixture was heated at 120°C for 30 minutes. The reaction was diluted with ethyl acetate, filtered, and concentrated. The crude product was subjected to silica gel chromatography (0-100% ethyl acetate/hexanes) followed by reverse phase HPLC (water/acetonitrile w/0.1% trifluoroacetic acid) to yield the desired product 3-([1,1'-biphenyl]-3-yl)-N-cyclohexylpyrrolidine-1-carboxamide (23.5 mg, 0.067 mmol). MS: m/z = 349.1 (M +1). ^{1}H NMR (400 MHz, DMSO-d_6) δ 7.70 – 7.62 (m, 2H), 7.58 – 7.28 (m, 7H), 5.75 (s, 1H), 3.74 (dd, *J* = 9.8, 7.5 Hz, 1H), 3.54 – 3.38 (m, 3H), 3.30 – 3.20 (m, 2H), 2.25 (dtd, *J* = 12.6, 6.5, 2.7 Hz, 1H), 2.00 (ddd, *J* = 20.2, 10.6, 8.7 Hz, 1H), 1.75 (d, *J* = 10.3 Hz, 2H), 1.68 (d, *J* = 12.0 Hz, 2H), 1.56 (d, *J* = 12.6 Hz, 1H), 1.31 – 0.99 (m, 5H) (Figure S7A).

Compound 2 (6-Isopropyl-3-(2-methoxybenzyl)-2-phenethylpyrimidin-4(3H)-one)

Compound 2 was obtained by reverse phase HPLC purification (water/acetonitrile w/0.1% trifluoroacetic acid)] of a historical sample (25mg) to yield the desired product 6-isopropyl-3-(2-methoxybenzyl)-2-phenethylpyrimidin-4(3H)-one (21.5 mg, 0.059 mmol). MS: m/z = 363.2 (M +1). ^{1}H NMR (500 MHz, Chloroform-*d*) δ 7.37 – 7.17 (m, 5H), 7.17 – 7.02 (m, 2H), 7.02 – 6.75 (m, 2H), 6.37 (s, 1H), 5.24 (s, 2H), 3.79 (d, *J* = 1.5 Hz, 3H), 3.20 (dd, *J* = 9.1, 6.9 Hz, 2H), 2.97 (m, 3H), 1.27 (dd, *J* = 6.8, 1.5 Hz, 6H) (Figure S7B).

Compound 2 can be synthesized by reacting 6-isopropyl-2-phenethylpyrimidin-4(1H)-one with 1-(bromomethyl)-2-methoxybenzene and potassium carbonate in acetonitrile and heating at elevated temperature. Purification by reverse phase HPLC (water/acetonitrile w/0.1% trifluoroacetic acid) yields the desired product.

Compound 3 (9-Benzyl-N-((1,2,3,4-tetrahydronaphthalen-1-yl)methyl)-9H-purin-6-amine)

9-Benzyl-6-chloro-9H-purine (50 mg, 0.204 mmol), (1,2,3,4-tetrahydronaphthalen-1-yl)methanamine oxalate (77 mg, 0.307 mmol) and triethylamine (85 µl, 0.613 mmol) were suspended in ethanol (1022 µl) and heated to 100°C overnight. The reaction mixture was concentrated, taken up in dichloromethane and purified by silica gel chromatography (0-50% (3:1) ethyl acetate/ethanol: hexane) to obtain the product 9-benzyl-N-((1,2,3,4-tetrahydronaphthalen-1-yl)methyl)-9H-purin-6-amine as a tan solid (63 mg, 0.171 mmol). MS: m/z = 370.2 (M +1). ^{1}H NMR (500 MHz, Chloroform-*d*) δ 8.45 (s, 1H), 7.71 (s, 1H), 7.42 – 7.22 (m, 7H), 7.13 (m, 3H), 5.87 (s, 1H), 5.37 (s, 2H), 4.02 (m, 1H), 3.81 (m, 1H), 3.22 (dt, *J* = 9.1, 4.8 Hz, 1H), 2.86 – 2.68 (m, 2H), 1.91 (m, 3H), 1.83 – 1.69 (m, 1H) (Figure S7C).

Reference:

1. Henneberger, C., Papouin, T., Oliet, S. H. R., and Rusakov, D. A. (2010) Long-term potentiation depends on release of D-serine from astrocytes, *Nature 463*, 232-U120.
2. Yang, C. R., and Svensson, K. A. (2008) Allosteric modulation of NMDA receptor via elevation of brain glycine and D-serine: The therapeutic potentials for schizophrenia, *Pharmacol Therapeut 120*, 317-332.
3. Rutter, A. R., Fradley, R. L., Garrett, E. M., Chapman, K. L., Lawrence, J. M., Rosahl, T. W., and Patel, S. (2007) Evidence from gene knockout studies implicates Asc-1 as the primary transporter mediating D-serine reuptake in the mouse CNS, *Eur J Neurosci 25*, 1757-1766.
4. Smith, S. M., Uslaner, J. M., Yao, L. H., Mullins, C. M., Surles, N. O., Huszar, S. L., McNaughton, C. H., Pascarella, D. M., Kandebo, M., Hinchliffe, R. M., Sparey, T., Brandon, N. J., Jones, B., Venkatraman, S., Young, M. B., Sachs, N., Jacobson, M. A., and Hutson, P. H. (2009) The Behavioral and Neurochemical Effects of a Novel D-Amino Acid Oxidase Inhibitor Compound 8 [4H-Thieno [3,2-b]pyrrole-5-carboxylic Acid] and D-Serine, *J Pharmacol Exp Ther 328*, 921-930.
5. Brown, J. M., Hunihan, L., Prack, M. M., Harden, D. G., Bronson, J., Dzierba, C. D., Gentles, R. G., Hendricson, A., Krause, R., Macor, J. E., and Westphal, R. S. (2014) In vitro Characterization of a small molecule inhibitor of the alanine serine cysteine transporter-1 (SLC7A10), *J Neurochem 129*, 275-283.
6. Sason, H., Billard, J. M., Smith, G. P., Safory, H., Neame, S., Kaplan, E., Rosenberg, D., Zubedat, S., Foltyn, V. N., Christoffersen, C. T., Bundgaard, C., Thomsen, C., Avital, A., Christensen, K. V., and Wolosker, H. (2016) Asc-1 Transporter Regulation of Synaptic Activity via the Tonic Release of d-Serine in the Forebrain, *Cerebral cortex*.
7. Sakimura, K., Nakao, K., Yoshikawa, M., Suzuki, M., and Kimura, H. (2016) A novel Na(+) - Independent alanine-serine-cysteine transporter 1 inhibitor inhibits both influx and efflux of D-Serine, *Journal of neuroscience research 94*, 888-895.
8. Fukasawa, Y., Segawa, H., Kim, J. Y., Chairoungdua, A., Kim, D. K., Matsuo, H., Cha, S. H., Endou, H., and Kanai, Y. (2000) Identification and characterization of a Na+-independent neutral amino acid transporter that associates with the 4F2 heavy chain and exhibits substrate selectivity for small neutral D- and L-amino acids, *J Biol Chem 275*, 9690-9698.

9. Nakauchi, J., Matsuo, H., Kim, D. K., Goto, A., Chairoungdua, A., Cha, S. H., Inatomi, J., Shiokawa, Y., Yamaguchi, K., Saito, I., Endou, H., and Kanai, Y. (2000) Cloning and characterization of a human brain Na+-independent transporter for small neutral amino acids that transports D-serine with high affinity, *Neurosci Lett 287*, 231-235.
10. Helboe, L., Egebjerg, J., Moller, M., and Thomsen, C. (2003) Distribution and pharmacology of alanine-serine-cysteine transporter 1 (asc-1) in rodent brain, *Eur J Neurosci 18*, 2227-2238.
11. Martineau, M., Baux, G., and Mothet, J. P. (2006) D-serine signalling in the brain: friend and foe, *Trends Neurosci 29*, 481-491.
12. Panatier, A., Theodosis, D. T., Mothet, J. P., Touquet, B., Pollegioni, L., Poulain, D. A., and Oliet, S. H. R. (2006) Glia-derived D-serine controls NMDA receptor activity and synaptic memory, *Cell 125*, 775-784.
13. Sun, D. Y., Jung, J., Rush, T. S., Xu, Z. W., Weber, M. J., Bobkova, E., Northrup, A., and Kariv, I. (2010) Efficient Identification of Novel Leads by Dynamic Focused Screening: PDK1 Case Study, *Comb Chem High T Scr 13*, 16-26.
14. Wassermann, A. M., Kutchukian, P. S., Lounkine, E., Luethi, T., Hamon, J., Bocker, M. T., Malik, H. A., Cowan-Jacob, S. W., and Glick, M. (2013) Efficient Search of Chemical Space: Navigating from Fragments to Structurally Diverse Chemotypes, *J Med Chem 56*, 8879-8891.
15. Paricharak, S., AP, I. J., Bender, A., and Nigsch, F. (2016) Analysis of Iterative Screening with Stepwise Compound Selection Based on Novartis In-house HTS Data, *ACS chemical biology 11*, 1255-1264.
16. Riniker, S., Wang, Y., Jenkins, J. L., and Landrum, G. A. (2014) Using information from historical high-throughput screens to predict active compounds, *J Chem Inf Model 54*, 1880-1891.
17. Rogers, D., and Hahn, M. (2010) Extended-Connectivity Fingerprints, *J Chem Inf Model 50*, 742-754.
18. Petrone, F. M., Simms, B., Nigsch, F., Lounkine, E., Kutchukian, P., Cornett, A., Deng, Z., Davies, J. W., Jenkins, J. L., and Glick, M. (2012) Rethinking Molecular Similarity: Comparing Compounds on the Basis of Biological Activity, *ACS chemical biology 7*, 1399-1409.
19. Brideau, C., Gunter, B., Pikounis, B., and Liaw, A. (2003) Improved statistical methods for hit selection in high-throughput screening, *J Biomol Screen 8*, 634-647.
20. Bemis, G. W., and Murcko, M. A. (1996) The properties of known drugs .1. Molecular frameworks, *J Med Chem 39*, 2887-2893.
21. Agrafiotis, D. K. (2003) Stochastic proximity embedding, *J Comput Chem 24*, 1215-1221.

Supporting Information Available: This material is available free of charge *via* the Internet.